DOCUMENT RESUME

ED 130 620 IR 004 151

AUTHOR Shettel, Harris H.; And Others

TITLE The Application of a By-Pass Technique to Programmed Instruction for Managerial Training.

INSTITUTION American Institutes for Research in the Behavioral Sciences, Pittsburgh, Pa.

SPONS AGENCY Westinghouse Electric Corp., Pittsburgh, Pa.
REPORT NO AIR-C-74-7/63-FR
PUB DATE Jul 63
NOTE 48p.; Archival document

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Comparative Analysis; Conventional Instruction; Educational Research; Individual Characteristics; Lecture; Management Education; *Programed Instruction; Retention; *Supervisory Training; Time Factors (Learning)

ABSTRACT

The feasibility of using programed instruction materials for a session of a Management Training Conference conducted for first-level supervisory personnel of the Westinghouse Electric Corporation was studied. A programing format utilizing a series of gated segments, each preceded by a test item, was designed to accommodate the variability among trainees in prior knowledge of the topic. The materials were administered to a group of 40 supervisors at a conference. A control group participated in the traditional lecture-discussion session led by an experienced instructor. Pre- and posttests were administered to both groups. While the lecture-discussion method was effective, the programed method was significantly better. The mean time required to complete the programed course was four hours and twenty minutes, compared to one and one-half hours required by the lecture. A retention test, administered four months later, showed no significant advantage for either group. The gating format of the materials proved effective in accommodating individual differences. (KB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED130620

AIR-C74-7/63-FR

THE APPLICATION OF A BY-PASS TECHNIQUE TO PROGRAMMED INSTRUCTION FOR MANAGERIAL TRAINING

Harris H. Shettel
Doris J. Clapp
David J. Klaus

American Institute for Research

Technical Assistance provided by

Robert Cadwallader
Westinghouse Electric Corporation

A Project Conducted under
Contract with

Westinghouse Electric Corporation

by

American Institute for Research
Pittsburgh, Pennsylvania

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

July 1963

IR 004 151

FOREWORD

The American Institute for Research is a private, nonprofit organization which carries out basic and applied investigations in the behavioral sciences and offers related technical services. The work of the Institute is entirely supported by grants and contracts from interested industrial organizations, government agencies, and foundations. Research projects, developmental efforts, and consulting services are carried out according to agreements between the Institute and sponsoring organizations.

The Institute has been vitally interested in the field of programmed instruction for the past several years. Within this rapidly developing area, the Institute has conducted both fundamental research studies and applied studies on the use of programmed instructional materials in industrial, military, and academic settings.

The project discussed in this report was conducted by members of the staff of the Training and Education Program and was under the general direction of Dr. David J. Klaus, Program Director. Project supervision was provided by Mr. Harris H. Shettel, Assistant Program Director. The programmed materials were prepared by Mrs. Doris J. Clapp.

Close liaison with the Westinghouse Electric Corporation was provided by Mr. Robert Cadwallader, Administrator, Manufacturing Supervisory Training.[1] Mr. Cadwallader was responsible for providing the project staff with an outline of the subject matter to be covered as well as with various other source documents. He provided verbal clarification of points in numerous conferences with the project staff and, in addition, reviewed program materials as they were being developed to insure their technical accuracy. Finally, Mr. Cadwallader prepared the criterion test materials and organized the field trials during which the programmed materials were administered to Westinghouse trainees. His assistance and support are gratefully acknowledged.

[1] Mr. Cadwallader is now Consultant, Professional Compensation Services.

TABLE OF CONTENTS

FOREWORD i

INTRODUCTION 1

PROGRAM OBJECTIVES, PROGRAM CONTENT, AND PROGRAM FORMAT 3

ADMINISTRATION OF FIELD TRYOUT 9

RESULTS OF FIELD TRYOUT
I. Effectiveness of the Programmed Materials and the Lecture Sessions . 13

II. Effectiveness of the gating format 23

DISCUSSION 33

SUMMARY 37

APPENDIX A - Outline of Westinghouse Employee Benefits Program

APPENDIX B - Programmed Group Raw Error Scores for Pretest and Posttest and Time to Complete Program

APPENDIX C - Lecture Group Raw Error Scores for Pretest and Posttest

APPENDIX D - Standard Error Scores (Z) and Percentile Scores (PS) for Programmed Group and Lecture Group (based on individual distributions)

APPENDIX E - Standard Error Scores (Z) and Percentile Scores (PS) for Programmed Group and Lecture Group (based on their combined distribution)

APPENDIX F - Raw Error Scores (ES), Percent Error Scores (PES), Standard Error Scores (Z), Retention Scores (RS), and Percentile Retention Scores (PRS) for Programmed and Lecture Group - based on the combined distribution of scores on a 4-month posttest

APPENDIX G - Data on Utilization of By-Pass Feature by Trainees

Figures

Figure 1 - Sample sequence showing gate frame, instructional frames, and criterion frame 6

Figure 2 - Percentile scores on pre- and posttests for both programmed and lecture groups based on their combined distribution 17

Figure 3 - Percentile improvement from pretest to posttest for both programmed and lecture groups based on their combined distribution 17

TABLE OF CONTENTS (cont.)

Tables

TABLE 1 - Number and Types of Frames Used in the Program 7

TABLE 2 - Average Raw Error Scores on Pre and Posttest for the Programmed Group 13

TABLE 3 - Within Group Comparisons Between Pretest and Posttest Standard Error Scores for Programmed Group 14

TABLE 4 - Average Raw Error Scores on Pre- and Posttest for the Lecture Group 15

TABLE 5 - Within Group Comparisons Between Pretest and Posttest Standard Error Scores for Lecture Group 15

TABLE 6 - Between Group Comparisons in the Error Reduction from Pretest to Posttest (Using Standard Error Scores Based on the Combined Distribution of the Two Groups) 16

TABLE 7 - Within Group Comparisons Between Initial Posttest and 4-month Posttest Percent Error Scores 19

TABLE 8 - Between Group Comparisons in the Number Correct on the 4-month Retention Test (Using Percentile Scores Indicating Loss in Retention, Based on the Combined Distribution of the Two Groups) 20

TABLE 9 - Within Group Comparisons Between Pretest and 4-month Posttest Percent Error Scores 21

TABLE 10 - CORRELATIONS - Gate usage as a method of accommodating for variations in subject matter knowledge 25

TABLE 11 - CORRELATION - Gate usage as a method of improving the overall efficiency of the program 29

TABLE 12 - CORRELATION - Whether tenure alone (vs. the gates) could be used as a measure of prior knowledge . . . . . 3[illegible]

INTRODUCTION

This report describes a study conducted by the American Institute for Research for the Westinghouse Electric Corporation. The primary objective of the study was to determine the feasibility of using programmed instruction materials as a replacement for a portion of a one-week Management Training Conference periodically conducted in Pittsburgh for first-level supervisory personnel. While the effectiveness of programmed instruction techniques has been demonstrated in a wide range of applications, Westinghouse wished to determine the contribution of these techniques when introduced into their conference-training setting. The subject matter selected for this pilot study was one of the 20 topics normally covered during the conference, the Westinghouse Employee Benefits Program.

Since the programmed materials were to be prepared as a possible replacement for the usual lecture/discussion session covering the topic of employee benefits, it was important to measure the relative effectivenss of the program and the traditional lecture/discussion method of instruction being used. Also, since long-term retention was a specific objective of the program, it was decided that the effectiveness of the two techniques would be compared not only immediately after training, but at some point following training as well. Appropriate pre- and posttests were prepared to permit these comparisons.

A factor which required special consideration in the development of the program was the nature of the subject matter especially with respect to the varying backgrounds of supervisory personnel who normally attend the conference. The Employee Benefits Program is of importance to all employees of Westinghouse. This is particularly true of supervisors, who may be called upon at any time to answer questions about features of the program put to them by hourly and salaried employees working under their direction. For this reason, various amounts of information concerning employee benefits ordinarily are accumulated by supervisors in the course of performing their jobs. The extent of such prior knowledge was expected to vary as a function of length of service with Westinghouse, their own particular interest in the various features of the program, the specific problems they have had to handle for employees, and so on.

For these reasons, it was quite likely that the knowledge the supervisory trainees had of the benefits program prior to their attendance at the training conference would vary considerably. In this case, a linear program which required all trainees to proceed through all of the programmed material would not appear to be the most efficient format. This project thus provided an opportunity to test techniques which would accommodate for individual differences in subject matter knowledge prior to training.

PROGRAM OBJECTIVES, PROGRAM CONTENT, AND PROGRAM FORMAT

The objective established for the programmed course was to teach the content of the Employee Benefits Program covered by an hour and a half long lecture/discussion period. A high level of recall was important so that supervisors would be able to accurately respond to inquiries from employees concerning benefit policies without resorting to available source documents. While no specific length of recall was specified, it was clear that it would have to be long term. In other words, the initial learning would have to insure the retention of knowledge over a period of many months or even years, depending partly on how much "on-the-job practice" a supervisor would receive subsequent to training. This point will be referred to again in the discussion section of this report.

Due to the anticipated variability among trainees in prior knowledge of the subject matter, it was necessary to employ a programming format which would allow those trainees with adequate knowledge of a given topic to skip corresponding instructional materials. Various possibilities were investigated in this connection, including intrinsic or "scrambled book" programming and various other branching formats. The essential problem was to discover a technique which would provide for (1) an accurate test of the trainee's knowledge prior to exposure to the instructional material, and (2) a method for skipping, or by-passing, unneeded materials which could be incorporated as an integral part of the program. The latter requirement was to permit the materials to be distributed and then used by the trainees on their own time.

An analysis of the content revealed that there were 53 separate topics to be covered. Each of these topics consisted of a discrete unit of information; knowledge concerning any one topic would not necessarily imply knowledge about any of the remaining topics. For example, a supervisor may know a good deal about eligibility for discount certificates without necessarily knowing anything about policies concerning reimbursement for moving expenses. Thus, each topic provided an appropriate opportunity to use a "test" item which would screen out those needing training on that topic from those who did not. The end result of this approach would be a series of "gated" segments. Any one segment would be taken by a trainee who did not pass the test item and would be by-passed by a trainee who answered the test item correctly.

The final format for the program consisted of 53 units of instructional material. Each unit was preceded by a gate item, or test item covering the subject matter presented in that particular unit. The gate item was superficially similar to a self-instructional frame, that is, it consisted of a short statement which required the trainee to make a response. It differed from a teaching frame, however, in that it was neither cued nor prompted, and required the trainee to make his response solely in the context of terminal or criterion conditions. This example of a gate frame taken from the program illustrates their general characteristics:

Gate 2

Mr. Briggs, an hourly employee, is laid off after 11 months' service; eight months later he is rehired. He will receive pay for the next holiday because _____

_____

_____.

This is designed to elicit fairly extensive knowledge about one feature of the holiday policy. It simulates a realistic context since a supervisor may be approached by an employee with a problem very similar to the one stated.

The trainees were instructed to write out their answer to each gate item in their own words, being careful to include all relevant information. They were cautioned that they should attempt to answer the gate frame only if they were fairly certain they could respond correctly. They were instructed that if they were at all uncertain about the answer, they should not attempt to guess but simply leave the gate frame blank and proceed through the unit of instruction immediately following. In the event that a trainee did respond to the gate frame, he was instructed to check his own answer with that given on the reverse side of the gate item and determine whether or not his response was correct. If, in his judgement, he was correct, he was told to skip the instructional frames and proceed to the next gate item. This was accomplished by simply turning a heavy divider page which separated the gated units. If the trainee did not answer the gate frame correctly (or, as noted, if he skipped it entirely), he was instructed to proceed through the instructional frames.

This procedure was considered to be a satisfactory solution to the problem of dealing with individual differences in subject matter knowledge existing prior to training.[1] The approach was likely to enhance the overall efficiency of the program since no one would be required to go through instructional materials covering areas in which he was already proficient. Of course, when a trainee made a response to a gate frame a certain amount of self-judgement on his part was required in determining whether or not he was correct. In writing the gate frames, however, every effort was ma'e to construct them so that the responses were as unambiguous as possible. In those cases where more than one point was to be stated, each one was listed separately so that they could be easily matched with the corresponding points made or not made by the trainee. Further control was exerted, in the preparation of gate items. They were designed to be both difficult and specific. This was likely to minimize guessing and therefore increase the reliability of the items as well as prevent the practicing of errors by trainees who were prone to guess.

The number of frames contained in each gated unit varied with the complexity of the topic covered. The gate frame shown above was followed by 17 instructional frames. Other gates contained as few as four instructional frames, and the average number for the entire program was nine. The last frame in each unit was a criterion frame, another cueless item of roughly the same difficulty level as the initial gate frame. Success on the criterion frame established at least temporary mastery of the topic corresponding to the level of mastery exhibited by trainees who were correct on the initial gate frame.

Each frame was printed on a separate small panel, approximately 2½ inches wide and 8½ inches long. These panels were bound overlapping each other so that only one frame at a time could be seen. The top frame on a page was always the gate frame and the instructional frames appeared in sequence directly below the gate. All of the frames for a given gate were contained on the equivalent of one or, in a few instances, two pages. A sample sequence from the program is shown in Figure 1. In the figure, all frames belonging to the unit are visible at one time which was not the case of the program in the form given to the trainees.

---

[1] This is, of course, not the only solution. For a discussion of other techniques and some of their advantages and disadvantages, see Shettel, H. H. Individual Differences in Subject Matter Knowledge and Programmed Instructional Format. Pittsburgh: American Institute for Research, January, 1963.

Gate 27

Mr. Cannon is hired by Westinghouse in 1952 at age 30. He accepts employment with another company in 1961. He does not have "vesting" in the Westinghouse pension plan because 1 _____ and also because 2 _____.

An employee leaving the Company before retirement will receive his accrued pension when he reaches age 65, if he has vesting under the plan. His age and length of service determines whether he has _____.

27-A

Employees must have at least 10 years' Credited Service with Westinghouse and be over age 40 when they leave to have vesting under the pension plan. An employee does not have _____ with less than _____ _____ or before age _____.

27-B

Mr. Harlow leaves the Company. He will have _____ in the pension plan if he has at least _____ Credited Service and is over age _____ when he leaves.

27-C

An employee who has left the Company will receive, at age 65, the amount accrued in his Westinghouse _____ plan if he was _____ _____.

27-D

Mr. Buckles, age 45, leaves the Company after five years' Credited Service. He has no vesting in the pension plan, because _____ _____.

27-E

Mr. Madison leaves the Company. He will have vesting in the pension plan if he is over age _____ and has at least _____ Credited Service.

27-F

Mr. Williams is leaving the Company. He will have vesting in the pension plan if _____.

27-G

Figure 1. Sample sequence showing gate frame, instructional frames, and criterion frame.

Notice that there are actually two points covered in the gated unit shown in Figure 1. The gate frame is constructed so that little room for doubt could exist concerning the accuracy of a response made to it. Also note that the last, or criterion frame, is essentially a restatement of the gate item.

The subject matter of the 53 topics covered in the program and the number of frames in each gated unit can be found in Appendix A.

In addition to the main body of the instructional materials, there was also a short introductory segment and a short review segment which all trainees were required to take. The introductory segment consisted of 19 frames and the review segment consisted of 49 frames. The number of instructional frames contained in the gated portion of the program totaled 475. The above figures are summarized in Table 1.

TABLE 1

Number and Types of Frames Used in the Program

Number of Gate Frames	53
Number of Instructional Frames (not including gate frames)	475
Average Frames/Gate (not including gate frame)	9
Number of Introductory Frames	19
Number of Review Frames	49
Total Frames in Program	596

The instructional frames within a gate were constructed according to the small step, overt response, linear programming technique. This approach was considered appropriate for subject matter with a high verbal content in which a large repertory of terms, facts, and concepts must be built up. Examples were extensively used both to add interest to the program and to supply concrete situations in which the trainees could practice applying the principle being covered. Answers were written directly in the program booklet by the trainee in the spaces provided (as shown in Figure 1). Each answer was immediately confirmed by having the trainee look at the correct response on the back of each frame.

The procedure followed in the tryout and revision of these materials included the successive administration of the program to individual subjects. The tryout subjects were brought in and administered the program under carefully monitored conditions. Error data and subject reactions were noted and revisions made in the program on the basis of this information. The error rate for the last five tryout subjects averaged between 5% and 6%.

ADMINISTRATION OF FIELD TRYOUT

Following their reproduction, the programmed instruction materials were administered to a group of 40 first-level supervisors attending the Management Training Conference during the month of February, 1962. This was the experimental group.

The control group for this study was a group of 36 supervisors attending a similar conference in April, 1962. For this group, instruction in the subject of employee benefits was given in a one and one-half hour lecture/discussion session by an experienced instructor.

For the programmed group, the session opened with a short orientation on programmed instruction. The trainees were told that they were participating in a study to determine the feasibility of using materials of this type in place of more conventional training methods. A pretest was then administered to measure the trainees' initial knowledge of the subject matter. After the pretest, the programmed materials were distributed. The men were asked to take the materials with them and to have them completed within a three-day period, i.e., the materials were passed out on Tuesday and the trainees were to report back for their final session on Thursday. There was no conference time allotted to work on the materials and, in view of the full conference schedule, it was expected that evenings would have to be used for this purpose. The trainees were told that, upon their return on Thursday, a posttest would be administered.

The control group was also given a pretest and, immediately following their instructional session, a posttest.

The test materials were prepared by Westinghouse and were based on the content of the programmed materials and the lecture/discussion session. Two parallel forms of the test were prepared, one containing 52 items and one containing 49 items. These forms of the test were used for both the experimental and control groups. Either form of the test could be used as a pretest and the other form as a posttest. There was approximately one multiple-choice test question for each of the gated segments of the program. The question on the 49 item test that covered "vesting" (see Figure 1), is as follows:

Mr. Stanley leaves the company. He has vesting rights in the Pension Plan if he is over the age of _____,

a. thirty-five
b. forty
c. forty-five
d. fifty

and has _____ years' Credited Service.

a. five
b. ten
c. fifteen
d. twenty

Note that the question is broken down into two parts as was the gate frame for this topic.

The two forms of the test were assigned randomly to both groups for the pretest. An individual who received the 49-item form as a pretest would be given the 52-item form as a posttest and vice versa. Due to a clerical error in the handling of the posttest booklets, most trainees in the experimental (programmed) group received the same test form for their posttest as they had taken as a pretest. While this could have adversely affected the validity of the study, the results (described in the next section) indicate this was not so.

The free time available to conference participants was limited. Most of the trainees in the programmed group reported that they had considerable difficulty completing the materials by the time they were due, being forced to utilize even a 15 or 20 minute free period to work on the program. Six men out of the initial group of 46 failed to complete the materials and were eliminated from the study. Conditions such as these are certainly less than optimal for program administration. However, they may realistically apply to any conference of this type where sufficient time <u>within</u> the conference schedule cannot be provided.

The men were asked to carefully record the time they spent on the program and an appropriate form for this purpose was contained in each program booklet. Due to the stop-start nature of program use, there are almost certain to be errors in the times reported. However, since there is no way of determining the direction of these errors, training time is presented here as reported by the trainees.

In interpreting the results of this study, it is important to realize the overall conditions under which the program was utilized. To review briefly, while the lecture/discussion session covered only one and one-half hours, the program was used over a three-day period. Thus, there was up to a three-day delay between the time some of the programmed materials were used by the trainee and the time when his knowledge of these materials was examined on the posttest. The programmed group was fully engaged during the day in other conference sessions dealing with other topics. While these conditions are less than ideal from an experimental point of view, they do provide data which reflect the result of program usage under realistic circumstances.

RESULTS OF FIELD TRYOUT

I. Effectiveness of the Programmed Materials and the Lecture Sessions

In analyzing the results of the field tryouts, several questions were posed regarding the amount learned:

A. Were the programmed materials an effective means of training?

B. Were the lecture sessions an effective means of training?

C. Was there a difference in the amount of learning from the programmed materials and the lecture session?

D. Was acquired knowledge retained over a four-month period?

E. Was there a difference in the amount of retention achieved with the programmed materials and the lecture sessions?

The results of the statistical tests of the data relating to the above questions will be discussed in the following sections.

A. The effectiveness of the programmed materials in terms of immediate learning.

If the programmed materials developed for this project were an effective means of training, there should have been a significant decrease in errors between the pre- and posttests. The average raw error scores on these tests for the programmed group are shown in Table 2. The average number of item errors on the pretest was 28.7 with a range from 15 to 39; the average number of item errors on the posttest was 9.5 with a range from 2 to 27.[1]

TABLE 2

Average Raw Error Scores on Pre- and Posttest for the Programmed Group

Pretest	Initial posttest	Error reduction from pretest
28.7	9.5	19.2

Due to the different number of items on the two test forms, it was necessary to convert the raw error scores noted above to standard (Z) scores (based on the standard deviation from the mean of the programmed group only) before statistically comparing pre- and posttest results. Table 3 shows the average pretest and posttest

[1] The error scores on the pretest and posttest for each trainee in the programmed group are shown in Appendix B.

standard scores and the average improvement in standard scores for the programmed group.[1] A comparison of these mean posttraining and pretraining test scores yields a t of 24.51 which is significant at the .01 level.[2]

TABLE 3

Within Group Comparisons Between Pretest and Posttest Standard Error Scores for Programmed Group

Pretest	Initial Posttest	Error Reduction from Pretest	Conclusion
0.8774	-0.8776	1.755 (p = .001)	Learning occurred in P group.

Before accepting this difference, it is necessary to consider whether the clerical error in distributing the tests prejudiced the results, that is, whether those who received different tests did poorer than those receiving the same test. Three of the 40 trainees received different tests. Did these individuals compare favorably in overall gain with the 37 trainees who received identical test forms? In terms of standard scores, the average improvement for the trainees taking the same test was 1.738; the average improvement for the trainees taking different tests was 1.963. Thus, the three trainees who received different test forms actually did better than the average for the remaining 37 trainees. There appears to be no evidence from this limited sample to support the notion that the clerical error in test distribution impaired the validity of the test results. It is, therefore, reasonably safe to conclude that the data reported above accurately reflect the actual improvement in achievement between pre- and posttest.

B. The effectiveness of the lecture sessions in terms of immediate learning.

In the course of a second Management Training Conference, the lecture/discussion format was used to teach trainees about the Employee Benefits Program. After the pretest was administered, approximately 1 hour 20 minutes was devoted to a lecture and discussion of employee benefits. The outline followed in this lecture was that

[1] Appendix D shows the standard score and the corresponding percentile score on both the pretest and the posttest for each trainee.

[2] A t of 2.58 would have met this level of significance. Therefore, the difference noted here is well beyond that required for statistical significance.

which was normally used for this topic and which had been used as the basis for the development of the programmed materials. Immediately following the lecture/discussion period, the posttest was administered. The pre- and posttests were the same as those used for the programmed group. The tests had been correctly distributed in this group so that each trainee's posttest was a different form than his pretest.

The number of trainees in this group completing both the pretest and the posttest was 36. There was no control over the composition of this group. As in the case of the programmed group, the men were first-level supervisors who had not previously attended such a conference.

The average error scores on the pre- and posttests for the lecture group are shown in Table 4. The average error score on the pretest for the lecture group was 25.9 with a range from 14 to 33, and the average error score on the posttest was 11.4 with a range from 5 to 23.[1]

TABLE 4

Average Raw Error Scores on Pre- and Posttest for the Lecture Group

Pretest	Initial posttest	Error reduction from pretest
25.9	11.4	14.5

The corresponding standard (Z) scores and the average improvement in standard scores are shown in Table 5. Again these standard scores were based on the mean and standard deviation of the lecture group only and were not combined with the programmed group.

TABLE 5

Within Group Comparisons Between Pretest and Posttest Standard Error Scores for Lecture Group

Pretest	Initial Posttest	Error reduction from Pretest	Conclusion
0.8789	-0.8789	1.758 ($p = .001$)	Learning occurred in L group.

[1] The error scores on the pretest and posttest for each trainee in the lecture group are shown in Appendix C.

The improvement in achievement for the lecture group is significant at the .01 level. Thus, the lecture method is also seen to result in significant gains in immediate achievement.

C. Comparing the programmed group and the lecture group.

Amount of learning: Of major interest is a comparison between the lecture improvement score and the improvement score obtained from the programmed material based on their combined distribution. The standard scores and the improvement scores for both groups are shown in Table 6.[1] The difference in the amount of improvement for the two groups was found to be significant at the .01 level.[2] Thus, while the lecture/discussion method was effective, there was a significant increase in the effectiveness of instruction in the group which had received the programmed materials.

TABLE 6

Between Group Comparison in the Error Reduction from Pretest to Posttest (Using Standard Error Scores Based on the Combined Distribution of the Two Groups)

Group	Pretest	Initial posttest	Error reduction from pretest	Conclusion
P	+1.0030	-.9630	1.966	
L	+0.7222	-.7669	1.489	
Diff.			0.477 (p=.001)	The groups differed significantly in the amount learned from pretest to posttest

These data are shown graphically in Figures 2 and 3. Figure 2 shows the combined percentile scores based on number of items correct on the pre- and post-

[1] While the lecture group started at a higher level (fewer errors), the correlations between pretest scores and gain scores for the two groups were not significantly different. Therefore, we may assume that the gain scores for the lecture group were not spuriously lower due to this ceiling effect.

[2] The standard scores and percentile scores for the programmed group and the lecture group, based on their combined distribution, are shown in Appendix E.

tests for both the lecture and programmed groups. Note that the programmed group started with poorer performance and ended with higher performance than did the lecture group. The difference in actual achievement between the two groups is shown in terms of combined percentile scores in Figure 3. The magnitude of this difference is on the order of 10% in favor of the programmed group.

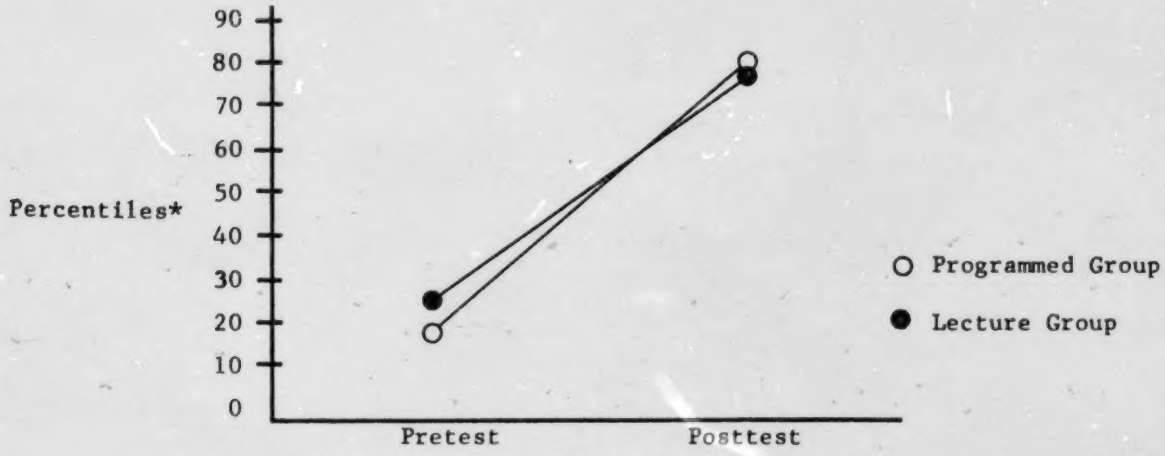

Fig. 2. Percentile scores on pre- and posttests for both programmed and lecture groups based on their combined distribution.

*Percentiles corresponding to area under the normal curve based on Z scores for combined distribution - programmed and lecture.

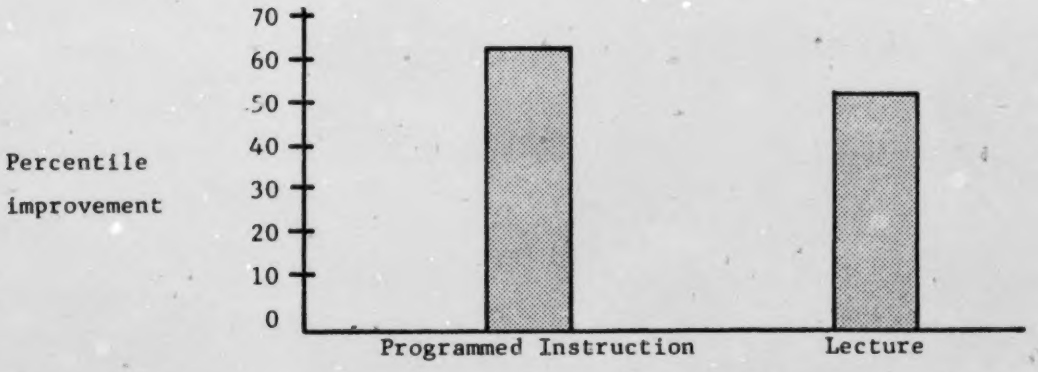

Fig. 3. Percentile improvement from pretest to posttest for both programmed and lecture groups based on their combined distribution.

Efficiency: The mean time to complete the programmed materials was 4 hours 20 minutes. The range was 1 hour 40 minutes to 8 hours 45 minutes.[1] In contrast, the lecture session lasted approximately one and one-half hours.

It is evident that the lecture method took considerably less time than the average time for those using the programmed materials, even allowing for possible errors in reporting the time for the programmed group. In answering the question, "Which is the best program?", one is faced with a trade-off problem of amount learned as against the time spent in training. Questions such as this can be adequately answered only by considering all important parameters, such as cost of training under the two methods, logistic advantages of one method over the other, cost of development, and so on. It could be argued, for example, that the economic advantage gained by sending the programmed materials out in the field, thus making it unnecessary to bring the men in to the home office for training, would more than compensate for the difference in training time. However, unless the entire conference is so programmed, this advantage would disappear, since the men would still have to be brought in for the other portions of the conference.

Other problems in interpreting the data should be noted. There was no control exercised over the selection of trainees for the two groups. Also, it is not possible to determine precisely what effect on performance resulted from the fact that the lecture group had less chance to forget information presented to them, i.e., tests for the lecture group were administered immediately after the one and one-half hour lecture session whereas, as previously noted, the programmed materials were used over a period of three days with a good deal of interpolated learning activity interspersed between the pre- and posttest. It is also not possible to measure the effect of the pretest on the kinds of questions asked by the trainees in the lecture/discussion session. It was pointed out by the leader of this session that the number of questions was unusually high and that some of them seemed to be directed specifically at areas that were covered in the pretest. This would, of course, tend to prejudice the results in favor of the lecture discussion method since the opportunity for directing questions at items covered in the pretest was not possible in the programmed group.

---

[1] Appendix B contains the reported time for each trainee in the programmed group.

D. Retention data for the programmed group and the lecture group.

Approximately four months after the administration of the posttest, a retention test was administered to both the programmed group and the lecture/discussion group.[1] Since it was necessary to mail the tests to the trainees at their respective locations, the tests were taken in an unsupervised, uncontrolled setting.

The initial posttest to 4-month posttest retention results indicate a considerable loss of knowledge for both groups. Table 7 shows the average percent error on the initial posttest and also on the 4-month posttest. While the increase in errors was not significant at the .05 level for either group, the increase in errors is pronounced, especially for the lecture/discussion group.

TABLE 7

Within Group Comparisons Between Initial Posttest and 4-Month Posttest Percent Error Scores

Group	Initial posttest	4-month posttest	Error increase	Conclusions
P	18.83	38.00	19.17 (t = 1.42)*	Loss in learning was not significant
L	22.61	43.04	20.43 (t = 1.80)*	Loss in learning approached significance

* $t > 1.96$ for $P = .05$

How did the two groups compare with respect to the increase in errors? Since the number of items on the 4-month retention test was different for the two groups, each raw error score on that test was converted to a percent error score by dividing by the number of items on the test (49 or 52). These percent error scores were then converted to standard error scores based on the combined distribution of the two groups. Next, these standard error scores were converted to retention

[1] The original 52-item test was given to the programmed group and the original 49-item test was given to the lecture/discussion group.

scores by subtracting the mean Z score on the initial posttest for that group (based on the combined distribution of the two groups).[1] The two groups were then compared in terms of percentile scores based on these converted retention scores.[2] The average percentile score (indicating relative loss in retention) for each group is shown in Table 8. A comparison of these scores yields a t of .253 which is not significant at the .05 level.

TABLE 8

Between Group Comparison in the Number Correct on the 4-Month Retention Test (Using Percentile Scores Indicating Loss in Retention, Based on the Combined Distribution of the Two Groups)

Group	Percentile loss in retention	Conclusion
P	68.509	
L	77.189	
Diff.	8.680 (t = .253)*	The two groups did not differ significantly in the amount retained from the initial posttest to the 4-month posttest

*t > 1.96 for P = .05

Thus, the superiority of the programmed group on the immediate posttest did not hold up under a 4-month period of retention. These data do not suggest at what point the posttraining scores ceased to be significantly different from each other, except that it did occur at some point before the end of the 4-month period.

[1] It was not possible to subtract the posttest score achieved by each individual since the retention test scores were not identified according to the individuals achieving these scores.

[2] All of these scores are shown for each trainee in Appendix F.

The fact that the loss in learning during the 4-month retention test was not significant for either group suggested one additional question: was there any difference between retention scores and pretest scores? Table 9 shows the average percent error on the pretest and on the 4-month posttest for both groups. The decrease in errors was not significant at the .05 level for either group. Thus, while the loss in learning was not significant, neither was the amount retained significantly different from knowledge on the pretest.

TABLE 9

Within Group Comparisons Between Pretest and 4-Month Posttest Percent Error Scores

Group	Pretest	4-Month posttest	Error reduction	Conclusions
P	56.88	38.00	18.88 ($t = 1.33$)*	Amount retained is not significantly different from knowledge on pretest
L	51.43	43.04	8.39 ($t = 0.70$)*	Amount retained is not significantly different from knowledge on pretest

*$t > 1.96$ for $p = .05$

RESULTS OF FIELD TRYOUT

II. Effectiveness of the gating format

A. The extent to which trainees utilized the gates

The comparative effectiveness of the programmed materials was only one area of interest in this study. Equally important was the effectiveness of the gating format described earlier. In other words, was the effort expended in developing the gated frames justified? Did the trainees make use of the gates, and if so, what effect did they have on achievement?

In answering these questions, a distinction must be made between gates which were attempted by the trainee as opposed to those that were answered correctly, resulting in a by-pass of instructional material. The total number of gates successfully answered by the 40 trainees was 408, or an average of about ten per trainee. By-passing ten out of the 53 units of instruction would save approximately 1/5 of the total maximum time and effort required to go through all units. The total number of gates attempted (right or wrong) was 1,525, for an average of 38 gates per trainee. Thus, while the average trainee attempted 38 gates, he only answered ten of them correctly, for a success ratio of about 1:4.

Overt practice of responses is one of the key elements of the programmed learning concept, and should, in principle, work as well for learning incorrect responses as it does for learning correct responses. Therefore, every effort was made in preparing gate items to reduce the amount of guessing; each gate item was made difficult enough so that only those who knew the answer would attempt it, but not so difficult that even those who knew the subject matter would be unable to answer it correctly. Since the format used could not prevent a trainee from guessing if he wished to do so, even though he knew nothing about the subject, it was impossible to achieve perfection. Having had no previous experience with this format, there was no standard against which to judge how well guessing was minimized. Future efforts using this format could, no doubt, reduce even further the number of error responses made by the trainees. (It should be noted that the preprogram instructions to the trainees were quite explicit in pointing out that guessing should not be attempted).

It was possible to shed some light on the seriousness of the guessing problem by correlating individual success ratios with decrease in test errors. If guessing was detrimental, one would predict a negative relationship (i.e. the greater the discrepancy between attempts and successes on the gates, the less will be the improvement in scores between the pre- and posttest). The actual correlation was -.04 (the standard error of the correlation was .16) which does not support the hypothesis that making errors is detrimental to learning. Problems in interpreting some of the gate responses made by the trainees (noted in the next section) and the general lack of experimental control over conditions in a feasibility study of this kind, suggest caution in interpreting these data.

Other indications of gate usage can be seen from the following data. At least one of the 53 gates was attempted by 39 of the 40 trainees, with only one trainee going through the entire program without ever attempting a single gate item. At least one unit was by-passed, that is, the gate item answered correctly and the instructional material skipped, by 35 of the 40 trainees. Thus, only five trainees failed to derive any time and effort advantage from the gate format.

Actual usage of the gates for each trainee can be seen in Appendix G, which provides data on (1) the number of gates unsuccessfully attempted, (2) the number of gates which were not attempted at all, and (3) the number of gates correctly answered. An examination of this table reveals some interesting characteristics concerning a few of the trainees. One trainee (#29) apparently attempted all of the gates with no correct answers at all.[1] Three trainees (#27, 38, 39) attempted 50 of the gates with a total of only six correct answers. Guessing apparently was a behavior at high strength for some trainees.

In summary, the initial assumption that the training population did have varying amounts of knowledge of the subject matter prior to the course appeared to be a correct one, since the number of gates correctly answered varied among the trainees. The specific effect that the use of the gates had on learning and training time can now be discussed.

---

[1] This is the same person who took 8 hours and 45 minutes to complete the program.

B. Gate usage as a method of accommodating for variations in subject matter knowledge.

With regard to the use of gates as a method of accommodating for variations in subject matter knowledge, the primary consideration was: were the gates effective in allowing subjects who knew the materials to by-pass the programmed units and in requiring subjects without knowledge of the materials to take the programmed units? To answer this question, the number of units by-passed were compared with a) the decrease in errors from pretest to posttest; b) pretest errors; c) the decrease in errors from pretest to posttest when pretest scores are equated; and d) the posttest error scores. The results of these comparisons are summarized in TABLE 10. We shall now consider each of these comparisons in more detail.

TABLE 10
CORRELATIONS

Gate usage as a method of accommodating for variations in subject matter knowledge

Correlation Between ⟶	and ⟶ (# Gate Usage of Units By-passed)	Conclusion	
Decrease in errors between pretest & posttest	$r = -.34$ $\sigma_r = .14$ ($p < .05$)	As by-passes decreased, the amount learned from the program increased.	Gates were effective in allowing Ss who already knew the material to by-pass the programmed units and requiring Ss without knowledge to take the programmed materials.
Pretest errors	$r = -.42$ $\sigma_r = .13$ ($p < .01$)	Use of the gates correlated with the amount of previous knowledge brought to the program.	
Decrease in errors with pretest errors equated	partial correl. $= -.15$* $\sigma_r = .16$	With pretest scores equated, use of the gates did not correlate with amount of improvement. Thus, the relationship between use of gates & amount of improvement is primarily based on the variations in pretest scores.	
Posttest error scores	$r = -.12$** $\sigma_r = .16$	Use of the gates did not correlate with the amount of knowledge at the completion of the program.	

*$r > .32$ for $p < .05$
**$r > .31$ for $p < .05$

... Use of gates and the amount learned from the program.

Since the gate questions and the test questions were designed to measure essentially the same knowledge, one would expect no change in test scores, if a hypothetical trainee answered all the gates correctly and thus by-passed all of the instructional material in these units. On the other hand, one would expect to find the greatest improvement in posttest scores over pretest scores as a result of missing all of the gates and thus taking all of the instructional material. For this reason, an overall negative correlation would be expected between the number of topics actually by-passed (gate usage) and the decrease in error scores between pretest and posttest. The actual correlation obtained between the number of topics by-passed and the decrease in error scores is -.34 (the standard error of the correlation is .14). This correlation is significantly different from zero at the .05 level. Thus, as an increasing number of topics were by-passed, the drop in error scores tended to decrease as anticipated. This result suggests that the gain in knowledge between pre- and posttest was directly related to the use of the by-pass feature.

While the relationship between the use of the by-pass feature and drop in errors is probably not due to chance, the magnitude of the relationship was not great. A number of factors which could not be controlled during the field tryout may be responsible for this. There is reason to believe, for example, that some of the trainees occasionally answered the gate correctly but did not skip the subsequent instructional frames as they were directed. Using an entirely new method of training, these trainees may have been more interested in seeing how the various topics were developed by the program than they were in skipping units. To the extent that this occurred, the magnitude of the relationship between drop in errors and gate usage would be reduced, of course, since these occurrences had to be counted as a unit taken rather than as a unit by-passed.

It was not possible to statistically factor out this effect since there is no way to determine from the gate item whether a trainee answered the gate correctly and then took the program unit or whether he took the program unit and then went back and filled in the gate frame correctly. While the instructions did not suggest going back and answering the gate after having completed the frames for that unit, they did not specifically caution against such a practice. Therefore, units that

contained correctly answered gates plus filled in frames for that unit, could only be interpreted as "missed the gate -- took the program unit." More careful control of trainee performance under classroom or laboratory conditions would have provided an opportunity to more closely monitor the behavior of the trainees in this respect. The extent to which trainees did take gate units, even though they knew the material, would adversely affect the overall efficiency of the program, as measured by the relationship between improvement scores and training time, since trainees were spending time going through material which could not be expected to improve their scores on the posttest.

... Use of the gates and the amount of previous knowledge.

A direct approach to measuring the relationship between previous knowledge and gate usage is to correlate pretest scores with the number of gates actually by-passed. The correlation noted above was between the reduction in error scores and the number of topics by-passed, which is a measure of the effectiveness of the program as an instructional unit rather than a measure of the effectiveness of the gates as a selection device for those who already know the subject matter. The extent to which this latter relationship existed would be shown by a negative correlation between errors on the pretest and the number of units actually by-passed, i.e., the more errors, the fewer by-passes. The obtained correlation was -.42 (σ_r = .13) showing that as errors on the pretest decreased, the number of units by-passed tended to increase. All of the comments made above pertaining to the difficulty in scoring by-pass usage applies to this correlation as well. It is still the case that the score for by-passed units is diluted by those who did not seem to make proper use of the format.

... Use of the gates and the amount learned from the program when pretest scores are equated.

Another approach to measuring the relationship between error scores and the use of gates is to consider what would happen if all trainees had achieved the same pretest score. Under these conditions, the same number of gates would be used by each trainee and the same amount of improvement would result for each trainee. This is the same as saying that the relationship between the number of units used and the amount of improvement is primarily due to variations in pretest error scores; if this variation were eliminated, differences in gate usage and improvement scores would also be eliminated. In order to test this expectation, a

partial correlation was computed which, in effect, gave all trainees the same pretest score and then measured the amount of improvement against the number of units by-passed. The obtained correlation was -.15 (σ_r = .16). This correlation cannot be attributed to other than chance factors. Thus, there is additional, indirect evidence that a trainee cannot expect to improve his score by going through a program unit if he has already answered the gate correctly. This indicates that the gate frames did a rather effective job of screening out those who did not need the instructional material from those who did need it insofar as meeting the terminal behaviors of this program was concerned (that is, answering the test items correctly).

... Use of the gates and the level of proficiency on the initial posttest.

Turning this problem around in the other direction, one would also predict that the same performance could be expected on the posttest by those trainees who by-passed units as would be found by those trainees not by-passing the units. That is, if the by-pass units were effective, they would tend to bring everyone up to the same level of proficiency. The gates can be considered in this sense as a leveler, differentially affecting an individual's knowledge depending on his previous knowledge, so that at the end of the program, everyone would score as well in areas where they were deficient as those who already knew the material. Thus, a zero correlation should exist between the use of the gates and error scores on the posttest. The obtained correlation between the use of the gates and error scores on the posttest was -.12 (σ_r = .16). A test of the hypothesis that the correlation is zero yields a t of only .75 which is not significantly different from zero at the .05 level. Thus, the performance on the posttest for those trainees who used the gates was not significantly different from the performance of those trainees who went through the instructional material.

In summary of the above correlations, it can be said that (1) the use of the gates correlated with the amount learned from the program; (2) the use of the gates correlated with the amount of previous knowledge brought to the program; (3) when pretest scores were equated, the use of the gates did not correlate with the amount learned; and (4) the use of the gates did not correlate with the amount of knowledge revealed at the completion of the program. All of these results support, in principle, the use of gating as a method of

accommodating for individual differences in subject matter knowledge.

C. Gate usage as a method of improving the overall efficiency of the program.

In addition to considering the gate as a method of accommodating for variations in subject matter knowledge, one would also expect that the gates would improve the overall efficiency of the program. Training time would be reduced for those trainees who, by virtue of previous knowledge, would not need to go through certain portions of the instructional materials. The sensitivity of the measurements in this area is diluted by two factors: (1) it is not likely that the actual times reported are precisely accurate, and (2) the units were not always utilized in the manner suggested in the instructions since, as noted, some trainees went through frames even though they had apparently answered the gate correctly. Nevertheless, training time is an important variable and the data available should be used despite these deficiencies.[1]

TABLE 11

CORRELATION

Gate usage as a method of improving the overall efficiency of the program

	Gate Usage (# of Units By-passed)	Conclusion
Training Time	$r = -.49$	
	$\sigma_r = .12$ ($p < .01$)	Overall program efficiency is positively influenced by the use of the by-pass format.

It would be expected that training time on the program would be decreased for those who made use of the by-pass feature of the gates. Thus, a negative correlation should be obtained between use of gates and training time. The actual obtained correlation, as shown in TABLE 11, was -.49 ($\sigma_r = .12$), which suggests that as use of the gates increased, training time decreased. This result supports the notion that overall program efficiency is positively influenced by

[1] In addition to possible errors in recording time, there is another element to consider in connection with training time. It has been noted that the trainees had to utilize small periods of time whenever they were available in order to complete the program on schedule. Since frequent stopping and starting is inefficient, this kind of intermittent use of the program would be expected to increase the overall time as compared with longer sessions with fewer interruptions.

the use of the by-pass format. It has already been determined that the use of the by-pass feature generally did not adversely affect posttest scores. Combining these two findings it can be said that, for the program under discussion, time was saved without loss of instructional effectiveness.

As noted, the mean time taken to complete the program was 4 hours 20 minutes. Taking the average figure for the number of units completed, which is 43 (10 being by-passed), and the average number of frames per unit, which is nine, the average trainee actually completed a total of 508 frames (9 x 43 = 387, plus the 53 gate frames, 19 introductory frames, and 49 review frames). Time spent on an individual frame is then seen to average approximately 1/2 minute. Since frame completion rate per hour for verbal material generally runs between 75 and 100 frames, it can be seen that trainees were progressing through this material at a slightly higher than "average" rate.

D. Tenure and the amount of previous knowledge.

While not directly related to the use of programmed materials, a point of general interest is the extent to which trainees' length of service as a supervisor would correlate with error scores on the pretest. As noted earlier, the trainees' prior knowledge of the subject matter was expected to vary partly as a function of tenure. Each trainee was asked to indicate on his test form his years in service as a supervisor, broken down into "less than one," "one to two," "two to three," and "more than three." Twenty of the 40 trainees had less than one year as a supervisor, fifteen had between one and two years, and the remaining five had two years or more. All of the latter trainees were combined into a "more than one year" group. The biserial correlation between length of service (less than one year or more than one year) and pretest score, as shown in TABLE 12, was -.18 (σ_{r_b} = .19). This is not significantly different from zero, and cannot be attributed to other than chance factors. Thus, tenure alone does not appear to be an accurate measure of the amount of knowledge one obtains concerning the Employee Benefits Program and could not in any way be used as a substitute for the gating principle employed in these materials.[1]

---

[1] This is an important point since it may be considered appropriate to use such external measures as length of service or kind of background experience as measures for screening out those who may not need certain portions of a programmed instructional course. These techniques are felt to be generally less adequate than techniques which require the trainee to actually produce the knowledge or behavior in question.

TABLE 12
CORRELATION
Whether tenure alone (vs. the gates) could be used as a measure of prior knowledge

	Length of service	Conclusion
Pretest errors	$r_{bis} = -.18^{*}$ $\sigma_{r_b} = .19^{*}$	Tenure alone does not appear to be an accurate measure of the amount of prior knowledge and could not be used as a substitute for the gating principle.

* $r > .31$ for $p < .05$

DISCUSSION

The programmed materials can only be called a "qualified" success, their length and their failure to provide for long-term retention being their two major shortcomings. An analysis of these problem areas will perhaps suggest ways in which they can be avoided in future, similar programming efforts.

A major factor contributing to the length of the program and to the less than adequate level of retention can be found by examining the objectives established for the program and their interaction with characteristics of the subject matter. As noted earlier, it was the goal of the program to provide sufficient training and drill so that supervisors could answer questions concerning the Employee Benefits Program without resorting to existing source documents or brochures. Furthermore, this knowledge was expected to remain at a high level of recall over an extended period of time.

In an effort to achieve this objective, the materials prepared included considerable drill and repetition of the many discrete facts that comprised the subject matter domain. It seemed clear that simply getting the trainee to make adequate responses within the program (low error rate) would not suffice. However, the amount of "overlearning" needed to meet the objective is difficult to specify since tryout subjects simply do not provide that kind of information. They indicate whether the program is teaching but do not make it possible to predict how much retention the program would produce. Furthermore, it was not clear how much practice on the job could be expected. Such practice would have a tendency to keep the knowledge acquired in the program at a high level of recall. But, even if a supervisor was asked one (different) question every week, which is unlikely, it would take approximately one year to cover all the topics. That would hardly represent an adequate level of practice.

It soon became apparent, as new drill frames were added, that the length of the program was going to exceed original estimates by a considerable margin, and that the time needed to take the program was going to exceed the time normally set aside for the conventional lecture presentation. Even so, the retention results were not encouraging, as noted in a previous section of the report. It appeared that on-the-job practice could not be counted upon to maintain the high level of knowledge exhibited at the completion of the program. The amount of drill and repetition contained in the program simply was not sufficient to compensate for this lack of practice on the job.

In the face of these results, two courses of action seem open. One is to revise the program so that the information is overlearned sufficiently to provide for long-term recall even under the low schedule of job practice. While the program that was prepared is no doubt capable of further "internal" improvement, it is believed that the most significant factor which would contribute to the improved recall of information would be to increase the amount of drill and repetition, i.e., to increase overall program length. This would be a costly solution, but it would perhaps be an acceptable cost if there were no suitable alternatives. However, there is one alternative that has important implications for the problem at hand.

This alternative would reach the same ultimate goal, but it would do so by a different path. Source documents currently exist which cover all the features of the Employee Benefits Program. Some are popularized publications for employees, other are official company manuals containing detailed discussions of all facets of the various plans. Many supervisors no doubt use one or more of these documents whenever they are confronted with a problem they cannot answer. The alternative being suggested here is to build the Employee Benefits training program around these documents rather than ignoring them. If present documents are not adequate, new ones could be prepared. Such documents would be characterized by the ease with which information could be found and by the clarity of the information presented. The training program would simply instruct supervisors in how to use these "job aids." Long-term "retention" of information would thus be insured by shifting the burden from the training program, which is transitory, to the source materials, which are permanent.

Program objectives are seen to have a dimension that may be overlooked in the initial program planning stages. This dimension is cost. It is not simply a question of deciding what <u>should</u> be taught and for how long it <u>should</u> be remembered. Equally relevant is the question of overall cost, made up of development cost and implementation cost. For the program under discussion, the objective initially established was not "wrong," nor was it impossible to meet, but it does appear that it is a relatively costly objective to meet. When costs go beyond a critical point (which can only be determined for each individual case by the user), alternatives (sometimes called compromises) are sought.

Training and education have always had problems of this kind. While programming offers no easy solutions, it does seem to force the problems into clearer relief where they can be objectively examined and, perhaps, more adequately solved. Like the boy who found the hole in the dike, programmed instructional techniques may actually appear to "cause" these problems. In point of fact, however, such tenchiques simply draw much needed attention to them. In the present instance, it was determined that neither programmed instruction nor the lecture/discussion session realized the intended objective of long-term retention of the subject matter. Bringing such data out into the open is certainly a first step in the re-examination of the interaction between training costs, training methods, and training objectives.

SUMMARY

The results of this feasibility study could be summarized as follows:

1. The programmed materials utilizing the gate concept taught a significant amount of information about the Employee Benefits Program even though these materials were used under less than ideal conditions.
2. As compared with a lecture/discussion session, the programmed materials showed themselves to be significantly more effective in their ability to teach as measured by immediate retention scores.
3. The group receiving the programmed materials required three times as long to complete them, on the average, than the time allotted to the lecture/discussion session.
4. There was no significant difference in retention of amount learned between the two groups when measured four months after training.
5. There was no significant difference between the level of knowledge exhibited four months after training and the level of knowledge exhibited on the pretest.
6. The gating principle was very effective in accommodating for individual subject matter differences. This conclusion is based on the following findings:
    A. The use of the gates correlated with the amount learned from the program.
    B. The use of the gates correlated with the amount of previous knowledge brought to the program.
    C. When pretest scores were equated, the use of the gates did not correlate with the amount learned.
    D. The use of the gates did not correlate with the amount of knowledge revealed at the completion of the program.
    E. The use of gates correlated with the time required to complete the program.

Thus, it is reasonable to conclude that the gating principle employed in this study successfully met its intended objective, i.e., it made it possible to accommodate for individual differences in subject matter knowledge with no apparent penalty in the teaching effectiveness of the material, and with average savings in time over an identical linear program on the order of 20%. This

improved efficiency over the linear format could be expected to apply to other programming efforts in which the subject matter of the program could be broken down into discrete elements and where substantial differences exist in program entrance skills.

APPENDIX A

	Frames in Each Unit
Introduction	19
Unit 1 - Holidays	5
Unit 2 - Eligibility requirements for holidays for hourly employees	18
Unit 3 - Eligibility requirements for holidays for salaried employees	7
Unit 4 - Review of eligibility requirements for holidays	4
Unit 5 - Observation of holidays falling on Saturday or Sunday	6
Unit 6 - Continued salary payments during sickness or disability - with less than one year's service - at the option of local management	11
Unit 7 - Continued salary payments during sickness or disability - with at least one year's service - at the option of local management	7
Unit 8 - Maximum total salary extension during all periods of sickness or disability within a 12-month period from date of last absence due to disability - at the option of local management	14
Unit 9 - Eligibility for discount certificates	6
Unit 10- Limit on number of discount certificates for major appliances; room air conditioners	11
Unit 11- Limit on number of discount certificates for TV sets	6
Unit 12- Limit on number of discount certificates for small appliances	14
Unit 13- Review of discount certificates	14
Unit 14- Employee stores	8
Unit 15- Procedure for purchase of products manufactured at other divisions	5
Unit 16- Eligibility requirements for vacations for all employees	14
Unit 17- Vacation periods for hourly employees	17
Unit 18- Vacation periods for salaried employees	16

	Frames in Each Unit
Unit 19 - Minimum eligibility requirement for a voluntary leave of absence - one year's service	7
Unit 20 - Voluntary leaves of absence with one to three years' service	7
Unit 21 - Voluntary leave of absence with four years' service and five or more years' service	8
Unit 22 - Military leaves of absence	4
Unit 23 - Annual encampment pay	12
Unit 24 - Vacations during annual encampment	5
Unit 25 - Employees serving on jury duty	7
Unit 26 - Requirements for normal retirement and early retirement under the pension plan for all employees	10
Unit 27 - Requirements for vesting under the pension plan for all employees	8
Unit 28 - Requirements for a disability pension	11
Unit 29 - Review on eligibility requirements for a normal retirement, early retirement, vesting, and a disability pension	11
Unit 30 - The amount of the monthly noncontributory pension	11
Unit 31 - Eligibility requirements for participation in the contributory pension plan	9
Unit 32 - The amount of the monthly contribution to the contributory pension plan	10
Unit 33 - The amount of the monthly contributory pension	14
Unit 34 - Moving expenses	8
Unit 35 - Introduction to the Westinghouse Insurance Plan	5
Unit 36 - Life Insurance, Nonoccupational Accidental Death or Dismemberment Insurance, and Travel Accident Insurance	19
Unit 37 - Workmen's Compensation	10

	Frames in Each Unit
Unit 38 - Time at which benefits from Nonoccupational Accident and Sickness Insurance will start	14
Unit 39 - Amount of benefits and period of coverage under Nonoccupational Accident and Sickness Insurance	11
Unit 40 - Review on Nonoccupational Accident and Sickness Insurance	11
Unit 41 - Coverage for hospital room expenses under Nonoccupational Hospital Expense Insurance	14
Unit 42 - Coverage for special hospital services under Nonoccupational Hospital Expense Insurance; Nonoccupational Surgical Operation Insurance	9
Unit 43 - Major Medical Protection for employees and dependents	12
Unit 44 - Maternity benefits	16
Unit 45 - Minimum service requirements for insurance coverages	13
Unit 46 - Hazards for which benefits are available under the Insurance coverages	14
Unit 47 - Invention award payments	10
Unit 48 - Suggestion system	13
Unit 49 - Employee stock plan: enrollment periods; time at which deductions start; time at which stock is purchased	14
Unit 50 - Employee stock plan - purchase price of stock	5
Unit 51 - Employee stock plan - amount of payroll deductions	7
Unit 52 - Periodic health examinations for supervisors	6
Unit 53 - Unemployment Compensation	8
Review	49

APPENDIX B

Programmed Group

Raw Error Scores for Pretest and Posttest and Time to Complete Program

Trainee Number:	Pretest Error Score		Posttest Error Score		Training Time Hours and Minutes
1	-26	(49)*	-2	(49)	3:48
2	-28	(52)	-3	(52)	4:06
3	-30	(52)	-3	(52)	3:40
4	-15	(49)	-4	(49)	3:33
5	-16	(52)	-4	(52)	2:58
6	-23	(49)	-4	(49)	4:48
7	-21	(52)	-5	(52)	2:30
8	-23	(49)	-5	(49)	1:40
9	-31	(52)	-5	(52)	4:09
10	-28	(52)	-6	(49)	3:34
11	-33	(52)	-6	(52)	3:45
12	-20	(49)	-7	(49)	5:23
13	-23	(49)	-7	(49)	4:20
14	-26	(49)	-7	(49)	4:20
15	-28	(49)	-7	(49)	4:11
16	-32	(52)	-7	(52)	4:37
17	-39	(52)	-7	(52)	3:52
18	-30	(52)	-8	(52)	3:42
19	-31	(52)	-8	(49)	6:27
20	-31	(52)	-8	(52)	3:33
21	-26	(49)	-9	(49)	3:31
22	-28	(49)	-9	(49)	4:38
23	-30	(49)	-9	(52)	4:45
24	-30	(49)	-9	(49)	4:49
25	-27	(49)	-10	(49)	5:18
26	-30	(52)	-10	(52)	3:12
27	-31	(49)	-10	(49)	6:02
28	-24	(49)	-11	(49)	2:56
29	-38	(52)	-12	(52)	8:45
30	-22	(49)	-13	(49)	3:09
31	-27	(52)	-13	(52)	2:15
32	-38	(49)	-13	(49)	4:20
33	-29	(52)	-14	(52)	4:50
34	-33	(52)	-14	(52)	4:31
35	-35	(49)	-15	(49)	6:53
36	-28	(52)	-16	(52)	5:50
37	-29	(49)	-16	(49)	5:59
38	-35	(49)	-18	(49)	5:05
39	-36	(49)	-18	(49)	6:10
40	-38	(52)	-27	(52)	3:41

* Numbers in parentheses indicate test version, i.e., 49-item test or 52-item test. As noted in the text, most trainees received the same test version on both the pretest and posttest.

APPENDIX C

Lecture Group

Raw Error Scores for Pretest and Posttest

Trainee Number	Pretest Error Score		Posttest Error Score	
1	-26	(49)*	-5	(52)
2	-27	(52)	-6	(49)
3	-26	(52)	-7	(52)
4	-17	(49)	-7	(52)
5	-29	(49)	-9	(52)
6	-19	(49)	-9	(52)
7	-25	(49)	-9	(52)
8	-22	(52)	-9	(49)
9	-22	(52)	-9	(49)
10	-26	(52)	-9	(49)
11	-24	(52)	-9	(49)
12	-14	(52)	-9	(49)
13	-31	(52)	-9	(49)
14	-28	(52)	-9	(49)
15	-26	(49)	-10	(52)
16	-23	(49)	-10	(52)
17	-27	(49)	-10	(52)
18	-25	(49)	-11	(52)
19	-25	(49)	-11	(52)
20	-21	(52)	-11	(49)
21	-26	(52)	-11	(49)
22	-25	(52)	-11	(49)
23	-30	(52)	-11	(49)
24	-25	(52)	-12	(49)
25	-33	(49)	-12	(52)
26	-25	(49)	-13	(52)
27	-31	(52)	-13	(49)
28	-28	(49)	-14	(52)
29	-29	(49)	-14	(52)
30	-32	(52)	-15	(49)
31	-28	(52)	-15	(49)
32	-25	(52)	-15	(49)
33	-27	(49)	-17	(52)
34	-33	(49)	-18	(52)
35	-25	(49)	-20	(52)
36	-29	(49)	-23	(52)

*Numbers in parentheses indicate test version, i.e., 49-item test or 52-item test.

APPENDIX D

Standard Error Scores (Z) and Percentile Scores (PS) for Programmed Group and Lecture Group (based on individual distributions)

Trainee Number	Programmed Group					Lecture Group				
	Pretest		Posttest		Difference	Pretest		Posttest		Difference
	Z	PS	Z	PS	Z	Z	PS	Z	PS	Z
1	0.760	.7764	-1.581	.0570	2.341	0.873	.8086	-1.687	.0458	2.560
2	0.674	.7498	-1.462	.0718	2.136	1.022	.8466	-1.516	.0648	2.538
3	0.845	.8009	-1.462	.0718	2.307	0.899	.8156	-1.441	.0748	2.340
4	-0.313	.3772	-1.386	.0829	1.073	-0.202	.4199	-1.441	.0748	1.239
5	-0.351	.3628	-1.376	.0844	1.025	1.231	.8909	-1.195	.1160	2.426
6	0.467	.6797	-1.386	.0829	1.853	0.037	.5148	-1.195	.1160	1.232
7	0.076	.5303	-1.291	.0983	1.367	0.754	.7746	-1.195	.1160	1.949
8	0.467	.6797	-1.289	.0987	1.756	0.406	.6576	-1.157	.1236	1.563
9	0.930	.8238	-1.291	.0983	2.221	0.406	.6576	-1.157	.1236	1.563
10	0.674	.7498	-1.191	.1168	1.865	0.899	.8156	-1.157	.1236	2.056
11	1.101	.8645	-1.205	.1141	2.306	0.653	.7432	-1.157	.1236	1.810
12	0.175	.5694	-1.094	.1370	1.269	-0.579	.2813	-1.157	.1236	0.578
13	0.467	.6797	-1.094	.1370	1.561	1.515	.9351	-1.157	.1236	2.672
14	0.760	.7764	-1.094	.1370	1.854	1.145	.8739	-1.157	.1236	2.302
15	0.955	.8302	-1.094	.1370	2.049	0.873	.8086	-1.071	.1421	1.944
16	1.016	.8452	-1.120	.1314	2.136	0.515	.6968	-1.071	.1421	1.586
17	1.614	.9467	-1.120	.1314	2.734	0.993	.8396	-1.071	.1421	2.064
18	0.845	.8009	-1.034	.1506	1.879	0.754	.7746	-0.948	.1716	1.702
19	0.930	.8238	-0.996	.1597	1.926	0.754	.7746	-0.948	.1716	1.702
20	0.930	.8238	-1.034	.1506	1.964	0.283	.6114	-0.918	.1793	1.201
21	0.760	.7764	-0.898	.1846	1.658	0.899	.8156	-0.918	.1793	1.817
22	0.955	.8302	-0.898	.1846	1.853	0.776	.7811	-0.918	.1793	1.694
23	1.150	.8749	-0.949	.1714	2.099	1.392	.9180	-0.918	.1793	2.310
24	1.150	.8749	-0.898	.1846	2.048	0.776	.7811	-0.799	.2122	1.575
25	0.857	.8043	-0.801	.2116	1.658	1.709	.9563	-0.825	.2047	2.534
26	0.845	.8009	-0.864	.1938	1.709	0.754	.7746	-0.702	.2414	1.456
27	1.248	.8940	-0.801	.2116	2.049	1.515	.9351	-0.680	.2483	2.195
28	0.565	.7140	-0.703	.2411	1.268	1.112	.8669	-0.579	.2813	1.691
29	1.528	.9367	-0.693	.2442	2.221	1.231	.8909	-0.579	.2813	1.810
30	0.370	.6443	-0.508	.3057	0.878	1.638	.9493	-0.441	.3296	2.079
31	0.589	.7221	-0.607	.2719	1.196	1.145	.8739	-0.441	.3296	1.586
32	1.931	.9733	-0.508	.3057	2.439	0.776	.7811	-0.441	.3296	1.217
33	0.759	.7761	-0.522	.3008	1.281	0.993	.8396	-0.209	.4172	1.202
34	1.101	.8645	-0.522	.3008	1.623	1.709	.9563	-0.086	.4657	1.795
35	1.638	.9493	-0.313	.3772	1.951	0.754	.7746	-0.160	.5636	0.594
36	0.674	.7498	-0.351	.3628	1.025	1.231	.8909	-0.530	.7019	0.701
37	1.053	.8538	-0.216	.4145	1.269					
38	1.638	.9493	-0.020	.4920	1.658					
39	1.735	.9586	-0.020	.4920	1.755					
40	1.528	.9367	+0.589	.7221	0.939					

APPENDIX E

Standard Error Scores (Z) and Percentile Scores (PS)

for Programmed Group and Lecture Group

(based on their combined distribution)

Trainee Number	Programmed Group				Lecture Group			
	Pretest		Posttest		Pretest		Posttest	
	Z	PS	Z	PS	Z	PS	Z	PS
1	0.801	.7884	-1.736	.0413	0.801	.7884	-1.427	.0768
2	0.851	.8026	-1.625	.0521	0.752	.7740	-1.313	.0946
3	1.049	.8529	-1.625	.0521	0.653	.7432	-1.229	.1095
4	-0.362	.3587	-1.525	.0636	-0.150	.4404	-1.229	.1095
5	-0.338	.3677	-1.526	.0635	1.119	.8684	-1.031	.1513
6	0.484	.6858	-1.525	.0636	0.061	.5243	-1.031	.1513
7	0.157	.5624	-1.427	.0768	0.696	.7568	-1.031	.1513
8	0.484	.6858	-1.419	.0780	0.256	.6010	-0.996	.1597
9	1.148	.8745	-1.427	.0768	0.256	.6010	-0.996	.1597
10	0.851	.8026	-1.313	.0946	0.653	.7432	-0.996	.1597
11	1.346	.9109	-1.328	.0921	0.455	.6754	-0.996	.1597
12	0.167	.5663	-1.207	.1137	-0.536	.2960	-0.996	.1597
13	0.484	.6858	-1.207	.1137	1.148	.8745	-0.996	.1597
14	0.801	.7884	-1.207	.1137	0.851	.8026	-0.996	.1597
15	1.013	.8445	-1.207	.1137	0.801	.7884	-0.932	.1757
16	1.247	.8938	-1.229	.1095	0.484	.6858	-0.932	.1757
17	1.940	.9738	-1.229	.1095	0.907	.8178	-0.932	.1757
18	1.049	.8529	-1.130	.1292	0.696	.7568	-0.833	.2025
19	1.148	.8745	-1.102	.1353	0.696	.7568	-0.833	.2025
20	1.148	.8745	-1.130	.1292	0.157	.5624	-0.785	.2162
21	0.801	.7884	-0.996	.1597	0.653	.7432	-0.785	.2162
22	1.013	.8445	-0.996	.1597	0.554	.7102	-0.785	.2162
23	1.224	.8896	-1.031	.1513	1.049	.8529	-0.785	.2162
24	1.224	.8896	-0.996	.1597	0.554	.7102	-0.679	.2486
25	0.907	.8178	-0.890	.1867	1.542	.9384	-0.734	.2315
26	1.049	.8529	-0.932	.1757	0.696	.7568	-0.635	.2627
27	1.330	.9082	-0.890	.1867	1.148	.8745	-0.573	.2833
28	0.590	.7224	-0.785	.2162	1.013	.8445	-0.536	.2960
29	1.841	.9672	-0.734	.2315	1.119	.8684	-0.536	.2960
30	0.379	.6476	-0.573	.2833	1.247	.8939	-0.362	.3587
31	0.752	.7740	-0.635	.2627	0.851	.8026	-0.362	.3587
32	2.070	.9808	-0.573	.2833	0.554	.7102	-0.362	.3587
33	0.950	.8289	-0.536	.2960	0.907	.8178	-0.239	.4056
34	1.346	.9109	-0.536	.2960	1.542	.9384	-0.140	.4443
35	1.753	.9602	-0.362	.3587	0.696	.7568	+0.058	.5231
36	0.851	.8026	-0.338	.3677	1.119	.8684	+0.356	.6391
37	1.119	.8684	-0.256	.3990				
38	1.753	.9602	-0.044	.4824				
39	1.859	.9685	-0.044	.4824				
40	1.841	.9672	+0.752	.7740				

APPENDIX F

Raw Error Scores (ES), Percent Error Scores (PES), Standard Error Scores (Z), Retention Scores (RS), and Percentile Retention Scores (PRS) for Programmed and Lecture Groups - based on the combined distribution of scores on a 4-month posttest*

Programmed Group					Lecture Group				
ES	PES	Z	RS	PRS	ES	PES	Z	RS	PRS
11	.2115	-2.089	-1.126	.1301	11	.2245	-1.949	-1.182	.1186
12	.2308	-1.882	-0.919	.1791	15	.3062	-1.070	-0.303	.3810
12	.2308	-1.882	-0.919	.1791	16	.3266	-0.850	-0.083	.4669
13	.2500	-1.675	-0.712	.2383	16	.3266	-0.850	-0.083	.4669
14	.2692	-1.468	-0.505	.3067	17	.3470	-0.631	+0.136	.5541
14	.2692	-1.468	-0.505	.3067	17	.3470	-0.631	+0.136	.5541
14	.2692	-1.468	-0.505	.3067	17	.3470	-0.631	+0.136	.5541
14	.2692	-1.468	-0.505	.3067	18	.3674	-0.411	+0.356	.6391
17	.3269	-0.847	+0.116	.5462	18	.3674	-0.411	+0.356	.6391
17	.3269	-0.847	+0.116	.5462	18	.3674	-0.411	+0.356	.6391
18	.3461	-0.640	+0.323	.6266	19	.3878	-0.192	+0.575	.7174
18	.3461	-0.640	+0.323	.6266	19	.3878	-0.192	+0.575	.7174
19	.3654	-0.433	+0.530	.7019	19	.3878	-0.192	+0.575	.7174
19	.3654	-0.433	+0.530	.7019	19	.3878	-0.192	+0.575	.7174
20	.3846	-0.226	+0.737	.7695	19	.3878	-0.192	+0.575	.7174
20	.3846	-0.226	+0.737	.7695	20	.4082	+0.028	+0.795	.7867
20	.3846	-0.226	+0.737	.7695	20	.4082	+0.028	+0.795	.7867
20	.3846	-0.226	+0.737	.7695	21	.4286	+0.248	+1.015	.8450
21	.4038	-0.019	+0.944	.8274	21	.4286	+0.248	+1.015	.8450
21	.4038	-0.019	+0.944	.8274	21	.4286	+0.248	+1.015	.8450
21	.4038	-0.019	+0.944	.8274	22	.4490	+0.467	+1.234	.8914
21	.4038	-0.019	+0.944	.8274	22	.4490	+0.467	+1.234	.8914
21	.4038	-0.019	+0.944	.8274	22	.4490	+0.467	+1.234	.8914
22	.4231	+0.188	+1.151	.8751	23	.4694	+0.687	+1.454	.9271
22	.4231	+0.188	+1.151	.8751	23	.4694	+0.687	+1.454	.9271
22	.4231	+0.188	+1.151	.8751	23	.4694	+0.687	+1.454	.9271
23	.4423	+0.395	+1.358	.9128	24	.4898	+0.906	+1.673	.9528
24	.4615	+0.602	+1.565	.9412	24	.4898	+0.906	+1.673	.9528
25	.4808	+0.809	+1.772	.9618	26	.5307	+1.347	+2.114	.9828
26	.5000	+1.016	+1.979	.9761	27	.5511	+1.566	+2.333	.9902
27	.5192	+1.223	+2.186	.9856	27	.5511	+1.566	+2.333	.9902
27	.5192	+1.223	+2.186	.9856	28	.5714	+1.785	+2.552	.9946
28	.5384	+1.429	+2.392	.9916	28	.5714	+1.785	+2.552	.9946
29	.5577	+1.637	+2.600	.9953	29	.5919	+2.005	+2.772	.9972
					29	.5919	+2.005	+2.772	.9972

*On the 4-month posttest the individual trainees were not identified. Therefore, trainee numbers cannot be presented here.

APPENDIX G

Utilization of By-Pass Feature by Trainees

(Total number of Gates = 53)

Trainee Number:	No. of Gates Attempted but Missed (Took Frames)	No. of Gates not Attempted (Took Frames)	No. of Gates Answered Correctly (Skipped Frames)
1	27	13	13
2	32	7	14
3	14	32	7
4	37	7	9
5	12	29	12
6	24	28	1
7	27	2	24
8	29	5	19
9	35	7	11
10	8	42	3
11	27	5	21
12	35	5	13
13	25	1	27
14	38	5	10
15	17	25	11
16	38	11	4
17	0	53	0
18	27	12	14
19	10	28	15
20	14	37	2
21	31	7	15
22	22	17	14
23	38	14	1
24	19	29	5
25	34	17	2
26	25	15	13
27	50	1	2
28	20	0	33
29	53	0	0
30	30	7	16
31	41	0	12
32	8	45	0
33	22	15	16
34	18	19	16
35	28	15	10
36	47	6	0
37	33	20	0
38	50	2	1
39	50	0	3
40	22	12	19